

Running Head: GIS AND DIABETES PREVENTION

Incorporating Geographic Information Systems into the Development of Diabetes

Programs in Ohio

Klara Box

The Ohio State University

College of Nursing

May 6, 2009

I. Introduction

Type 2 diabetes mellitus (DM) is a chronic disease with far reaching social and economic consequences. Individuals with DM are at heightened risk for atherosclerotic disease, coronary artery disease, stroke, peripheral vascular disease, blindness and infection (Huether & McCance, 2004). In addition, DM is the most common cause of end-stage renal disease (Huether & McCance, 2004). According to the Ohio Department of Health (ODH) DM is the sixth-leading cause of death in Ohio and Ohio's diabetes rate of 9.5% is higher than the national average of 8% (ODH, 2008). Furthermore, between 1995 and 2007 the prevalence of diabetes in Ohio increased 126%, from 4.2% to 9.5% of Ohio adults having been diagnosed with the disease (ODH, 2008). In 2007, the direct (medical care) and indirect (lost productivity) costs of diabetes in Ohio was an estimated \$5.9 billion (ODH, 2008). The high incidence, increasing prevalence, and enormous cost of diabetes in Ohio calls for diabetes programs to stop, delay and manage the onset of disease.

Intervention programs can prevent or delay the onset of type 2 diabetes mellitus (Pronk, Boucher, Jeffery, Sherwood & Boyle, 2004) and the need for such programs in Ohio is evident. An important component of a successful intervention program is to know *where* to implement it (Gesler, Hayes, Arcury, Skelly, Nash & Soward, 2004).

Geographic Information Systems (GIS) can be utilized to identify where prevention programs are needed most. GIS is a software tool that allows for geocoding of data (Riner, Cunningham & Johnson, 2004). The data inputted into GIS allows the user to map data in a way that displays spatial information in a striking and easy-to-read format (Gesler, et al., 2004). For example, GIS can identify which Ohio counties have a high prevalence of diabetes but few provider services allowing public health services to

allocate resources efficiently. Although GIS has been used to map where programs are needed in other states, it has not been utilized in Ohio to assist in developing diabetes programs. Thus, the purpose of this study is to geocode data for use in GIS in order to propose which Ohio counties are at highest risk and therefore in most need of diabetes programs.

II. Review of Literature

A. Type 2 Diabetes Mellitus

Significant factors in the onset of Type 2 diabetes mellitus (DM) are cellular resistance to insulin and a decrease in pancreatic production of beta cells with obesity being the single most influential disease risk factor (Heuther & McCance, 2004). The incidence of DM is rising nationally and in Ohio (ODH, 2008) and persons with DM are at heightened risk for secondary health problems, such as heart disease, kidney failure, stroke and blindness (Heuther & McCance, 2004). Health programs targeting DM are important in preventing complications associated with the disease (Pronk, Boucher, Jeffrey, Sherwood & Boyle, 2004) and health programs are particularly beneficial in areas of health disparity (Coberley, Puckrein, Dobbs, McGinnis, Coberley & Shurney, 2007; Davidson, 2004). Ohio's DM rate is higher than the national average and rural areas are especially at risk (ODH, 2008). Identifying where DM programs are needed most in Ohio is key to program success and a useful tool toward that goal is Geographic Information Systems (GIS) (Caley, 2004).

B. GIS as a Tool for Developing Diabetes Prevention Programs

GIS is computer software that is used to collect, store, retrieve and statistically interpret geographic or location-based information (Croner, Sperling, & Broome, 1996).

GIS databases include both spatial and nonspatial data (Croner, et al., 1996).

Examples of nonspatial data include demographic and socioeconomic indicators that are identified within geographic boundaries, such as cities and counties. Spatial data, by contrast, are exact geographical locations assigned using geocoding, also known as address matching (Croner, et al., 1996). Address matching converts an address, or set of coordinates, to a specific point on a map. The resulting points (or lines or areas) become a thematic layer of location-related data. Overlaying multiple themes on a single map has the potential to illustrate spatial relationships between data (Choi, Afzal, & Sattler, 2006; Riner et al., 2004; Gesler, 2004).

Several studies have found that GIS is an effective tool for designing and implementing health intervention programs (Caley, 2004; Coberley, et al., 2007; Riner et al., 2004; Gesler, 2004). Moss and Schell (2004) have observed increases in the health care application of GIS. Health care applications of GIS include tracking disease trends and describing where healthcare services are in relation to the populations that need them (Moss & Schell, 2004). Specific benefits of GIS include allowing practitioners to identify spatial relationships, patterns and trends that might otherwise go unnoticed and to prioritize the use of limited resources for improving public health (Riner, et al 2004).

Of particular significance is Coberley et al.'s (2007) study in which GIS was used to identify U.S. regions with high diabetes prevalence rates and to identify areas with health disparity. The authors found that certain U.S. regions are associated with poorer health outcomes and that health disparate individuals with diabetes have increased complications and mortality. The study concludes that targeting disease management programs in areas of health disparity helps to reduce gaps in care (Coberley, et al.,

2007). No studies were identified that have used GIS in the development or implementation of diabetes programs in Ohio.

B. Prevalence and Distribution of Diabetes Across Ohio.

Between 1995 and 2007 the prevalence of diabetes in Ohio rose 126% (from 4.2% to 9.5%) and Ohio's diabetes rate of 9.5% is higher than the national average of 8% (ODH, 2008). Prevalence of diabetes by county ranged from a low of 6.5% in Delaware, Fairfield, Knox, Licking, Madison, Morrow, Pickaway, and Union counties (interestingly these counties representing the lowest prevalence have a rate significantly higher than the 1995 overall Ohio prevalence rate of 4.2%) to a high of 11% and 11.1% in Lucas and Mahoning counties respectively. The mortality rates among Ohio's 88 counties ranged from a high of 52 per 100,000 residents, in Van Wert County, to a low of 13 per 100,000 residents in Carroll and Paulding counties with a state average of 31 per 100,000 persons (ODH, 2008). According to ODH (2008) mortality rates do not follow a set geographical pattern and that Ohions died from diabetes regardless of whether they lived in an urban, rural or suburban county.

Socioeconomic status, which is a type of nonspatial data (Croner, et al., 1996), shows that in Ohio diabetes is more common in households with an annual income of less than \$25,000 and that prevalence of DM increases as a person's education level decreases (ODH, 2004). In Ohio, therefore, lower socio-economic status is associated with DM (ODH, 2004). Risk factors for DM can be combined in GIS and displayed as maps to help target communities that would benefit most from prevention programs.

C. Diabetes Programs

Primary prevention of DM focuses on reducing the incidence of new cases through prevention or delay of disease onset (Pronk, Boucher, Jeffrey, Sherwood & Boyle,

2004). Prevention takes the form of pharmacologic or lifestyle interventions with lifestyle intervention being the most effective (Pronk, et al., 2004). The first step in an effective program is screening in order to identify individuals at high-risk (Pronk, et al., 2004). Risk indicators for DM include age, body mass index, waist circumference, hypertension, high blood glucose, and physical inactivity (Pronk, et al., 2004; Davidson, 2004; Signorello, Schlundt, Cohen, Steinwandel, Buchowski, McLaughlin, Hargraves & Blot 2007).

Based on a review of the literature Pronk et al (2004) identify key lifestyle interventions to prevent DM. These lifestyle modifications include weight loss of 7% of total body weight, dietary changes to include more fruits and vegetables and an increase in physical activity levels amounting to 150 minutes per week of moderate intensity exercise. According to Pronk, et al. (2004) these lifestyle changes are effective in preventing the onset of new DM disease and are even more effective than pharmacologic interventions. Those that receive the greatest benefit from diabetes programs are those that live in areas of health disparity (Coberley, et al., 2007; Davidson, 2004).

D. Summary

DM is a chronic disease with far reaching health implications. In Ohio the prevalence of DM is rising and continues to remain above the national average. The distribution of DM in Ohio is related to socioeconomic factors, such as poverty. Intervention programs promoting screening and lifestyle interventions are effective in DM prevention and in improving control in those with diagnosed disease. GIS is a tool increasingly being used by health care providers to develop health programs. GIS can

be utilized to help identify high-risk counties in Ohio that would most benefit from programs targeting type 2 diabetes mellitus.

III. Methodology

Following a review of the literature data for use in this project was collected from the Ohio Department of Health, the U.S. Census Bureau, the Centers for Disease Control and Prevention, and Melissa Data. The collected data was imported into GIS in order to generate maps and then analyzed using statistics.

A. Sample

The sample for this study includes all 88 Ohio counties. The study mapped the following four county-level measures.

B. Measures

1. Prevalence: The Centers for Disease Control (CDC) and Prevention National Diabetes Surveillance System provided estimated percentage of adults in each Ohio county with DM (CDC, 2005). Estimates of prevalence were established using data from the CDC's Behavioral Risk Factor Surveillance System (BRFSS) and data from the U.S. Census Bureau's Population Estimate Program. The BRFSS is an ongoing, monthly state-based telephone survey of the adult population, defined as 20 years of age or older (CDC, 2005). If respondents answer "yes" to the question, "has a doctor ever told you that you have diabetes?" then they are classified as having diabetes. Women who answer that they only had diabetes during pregnancy are not considered to have the disease (CDC, 2005).

Three years of data were used to improve the precision of yearly county-level diabetes estimates. For example, data from 2004, 2005, and 2006 were used for the

2005 estimate (CDC, 2005). Estimates were restricted to 20 years of age and older to be consistent with population estimates from the U.S. Census Bureau (CDC, 2005). County level prevalence estimates are based on indirect model-dependent estimates (CDC, 2005). The model-dependent technique uses a statistical model that “borrows strength” in making an estimate for one county from BRFSS data collected in other counties (CDC, 2005). Estimates were obtained using the Bayesian multilevel modeling approach. Separate models were developed for each of the four census regions: Northeast, South, West and Midwest (CDC, 2005).

2. Mortality: Ohio Department of Health county-level data was used to determine annual age-adjusted mortality rates of diabetes, per 100,000 persons, by county (ODH, 2008). The ODH Office of Vital Statistics identified diabetes deaths between 1995 and 2005 via a population-based, computerized database. Any person who had diabetes listed on their death certificate as an underlying cause of death was included in the analysis (ODH, 2008). Average annual direct age-adjusted rates were calculated for all Ohio counties for the five-year period between 2001-2005. Five years of data were combined to obtain a large enough sample for analysis because some Ohio counties have a relatively small number of residents and therefore a small number of diabetes deaths (ODH, 2008). ODH analysis of data was completed using SAS 9.1 (ODH, 2008).

3. Poverty: The U.S. Census Bureau provided data on the percent of all people living in poverty per Ohio county (US Census Bureau, 2005). The U.S. Census Bureau used data from the American Community Survey (ACS) as the basis for their county-level 2005 Small Area Income and Poverty Estimates (SAIPE) for all U.S. counties or county equivalents (e.g. parish, borough).

The model is multiplicative, meaning that the number of people in poverty is a series of predictors that are numbers, not rates, and the unknown errors are modeled (US Census Bureau, 2005). In order to estimate coefficients in the model, the U.S. Census Bureau took logarithms of the dependent variable (total number of people in poverty in each county as measured by ACS) and all of the independent variables. The independent variables include: the number of tax return exemptions, of all ages, on returns whose adjusted gross income falls below the official poverty threshold; number of food stamp recipients in July of the previous year; estimated total resident population as of July 1; total number of tax return exemptions; and Census 2000 estimate of the total number of people in poverty (US Census Bureau, 2005).

Taking the log of each of these variables was done because the distribution of the number in poverty has a large range- from zero in some U.S. counties to more than a million in the largest county- so that the distribution is highly skewed. Taking the logarithm of all variables centers the distribution and diminishes the influential effect of large counties on coefficient estimates.

4. Providers: Diabetes-related health care providers were retrieved from Melissa Data Corporation, a private firm based in California. Melissa Data is a database of addresses recognized by the U.S. Postal Service that is updated several times a year. Using the SIC code lookup feature in the Melissa Data database the number and location of all offices and medical doctors in Ohio were identified and then sorted to display diabetes-related health care providers only. Diabetes-related health care providers were defined for the purpose of this study as internal medicine practitioners, endocrinologists, internal medicine physicians/surgeons, medical centers, clinics operated by physicians, freestanding emergency medical clinics, primary care medical

clinic, diabetes specialist physician/surgeon, ophthalmologists, and general and family practice physicians/surgeons (Melissa Data, n.d.).

The resulting list of diabetes-related health care providers in Ohio included their locations in the form of an address. These addresses were converted into latitude/longitude coordinates using the Melissa Data feature GeoCoder Object® so that the location of providers could be displayed as point data on GIS generated maps. For statistical analysis provider data was also translated in GIS from latitude/longitude positions into their corresponding counties in order to calculate diabetes-related providers per 100,000 people by county.

C. Statistics

Statistical methods were used to identify high-risk counties and correlations between variables. High-risk counties were identified as those that concurrently have high prevalence, high increase in prevalence over time, and high mortality. Specifically, counties with high prevalence were defined as those in the top quartile of prevalence amongst all Ohio counties; counties with high mortality were defined as those in the top quartile of mortality amongst all Ohio counties; and counties with high increase in prevalence over time were defined as those in the top tertile of change amongst all Ohio counties.

Each Ohio county was assigned a number corresponding to the quartile of prevalence, the quartile of mortality and the tertile of change they rank in compared to other Ohio counties. For example, a county in the top quartile of mortality was assigned the number 4, a county in the third quartile of change a “3” and so on. In terms of change in prevalence counties were assigned numbers according to which tertile of change they fell under, a “3” being a county with the highest tertile of change, while a

“1” represents counties with the lowest amount of change. The sum of each county’s numeric coding was totaled in order to identify their overall risk. The highest risk counties were defined as those that after summing totaled a 10 or 11. A score of “11” is the highest possible score, meaning a county at highest risk, because in each category they were in the top quartile or top terile of risk compared to other Ohio counties (eg. $4+4+3=11$). Lower scores (3 being the lowest possible, eg $1+1+1=3$) represent counties that are at lower risk.

D. Procedures

Data collected on measures of prevalence, mortality, poverty and providers was compiled into an Excel spreadsheet. Spreadsheet variables were exported to ArcView GIS ArcMap 9.2 and overlaid on a map displaying prevalence of diabetes by county. Prevalence of diabetes is overlaid on a county level TIGER/Line Shapefile® map of Ohio, creating a choropleth (or thematic) map. Prevalence rates are grouped in five equal intervals and presented as a graded color series, where darker colors, in this case darker shades of green, represent higher diabetes prevalence, lighter shades of green represent lower diabetes rates. Change in prevalence was calculated by the difference in diabetes prevalence between 2000 and 2005 per county and displayed as a separate choropleth map where the dark red counties represent higher rates of change than the light red counties.

Percent of all people in poverty by county was overlaid on the diabetes prevalence map. Poverty rates are displayed as proportionally scaled circles. ArcMap created poverty classes with the natural breaks feature. This feature identifies class breaks that best groups similar values and maximizes the differences between classes.

Providers were defined as clinics and doctors that manage diabetes patients, for example diabetes specialists, endocrinologists and primary care medical clinics. Location of providers was displayed as point data based on latitude and longitude (lat/long) and overlaid on the diabetes prevalence map. For statistical analysis provider data was geocoded to convert lat/long positions into corresponding counties in order to calculate diabetes-related providers per 100,000 people by county.

Diabetes associated mortality rates per county are displayed as proportionally scaled circles that are overlaid on the diabetes prevalence map. ArcMap created mortality classes with the natural breaks feature. This feature identifies class breaks that best groups similar values and maximizes the differences between classes. Each map created includes a legend that makes the color shading scheme and symbols easy to interpret.

In addition to mapping the variables, statistical analysis was done using SAS 9.1 in order to identify correlations between variables. A Pearson correlation coefficient (r) was run in SAS in order to identify correlations between variables (Table 1).

IV. Results

The maps generated to illustrate the change in diabetes prevalence between 2000 and 2005 indicate that diabetes is increasing in all 88 Ohio counties (Figure 1). Figure 1 shows that counties around Lake Erie and areas of eastern Ohio had the greatest increases in prevalence. The map of poverty overlaid on prevalence indicates increasing prevalence is aligned with poverty (Figure 2). The map depicting the location of diabetes related providers shows that eastern counties have few providers (Figure 3). The map depicting mortality overlaid on prevalence suggests mortality follows patterns of poverty and prevalence (Figure 4).

Observations were confirmed using statistical analysis to identify counties with greatest cumulative risk- those counties in the top quartile of prevalence; top quartile of mortality and top tertile of change. Two counties were dropped from the analysis as outliers based on poverty data. From the summed risk score analysis two clusters of high-risk counties emerged. From a total of 10 Ohio counties that had a sum score of 10 or 11, eg, the highest risk counties based on prevalence, rate of change and mortality, five counties form a cluster along the State's eastern central edge along the Appalachian river (Jefferson, Belmont, Guernsey, Monroe and Washington counties). A second cluster, consisting of two counties, emerged in the southwest (Montgomery and Clark counties). The remaining three high-risk counties (Nan Wert, Huron and Wayne) are dispersed.

Pearson's r was significant for mortality and poverty ($r=0.29$), meaning there is an observed relationship between higher poverty rates and increased mortality from diabetes. No other correlations were found to be significant (for example there is not an observed relationship between poverty and number of providers).

V. Conclusion

GIS is useful in displaying health data in an easy-to-read map format. Maps can be useful starting points for public health nurses to utilize when developing intervention programs. Maps shows a large amount of information that is easy to read and can be used to help make decisions about resource allocation (Caley, 2004), such as were DM programs are needed most in Ohio.

The maps generated in this project illustrate two clusters of Ohio counties that are high risk and therefore most likely to benefit from diabetes intervention programs. One limitation of this study was the few numbers of variables included for analysis.

Future study would benefit from incorporating more DM risk variables such as race/ethnicity, education level and obesity.

Table 1- Pearson correlation coefficients of DM risk factors

	Providers (per 100,000)	People in poverty (%)	Mortality (per 100,000)
Providers (per 100,000)		0.1372	-0.04183
People in poverty (%)	0.13792		0.29096
Mortality (per 100,000)	- 0.04183	0.29096	
Change in prevalence (2000 - 2005)	0.13225	0.15925	0.02100

Figure 1 – Map of change in DM prevalence in Ohio counties

Change in Diabetes Prevalence in Ohio, 2000 - 2005

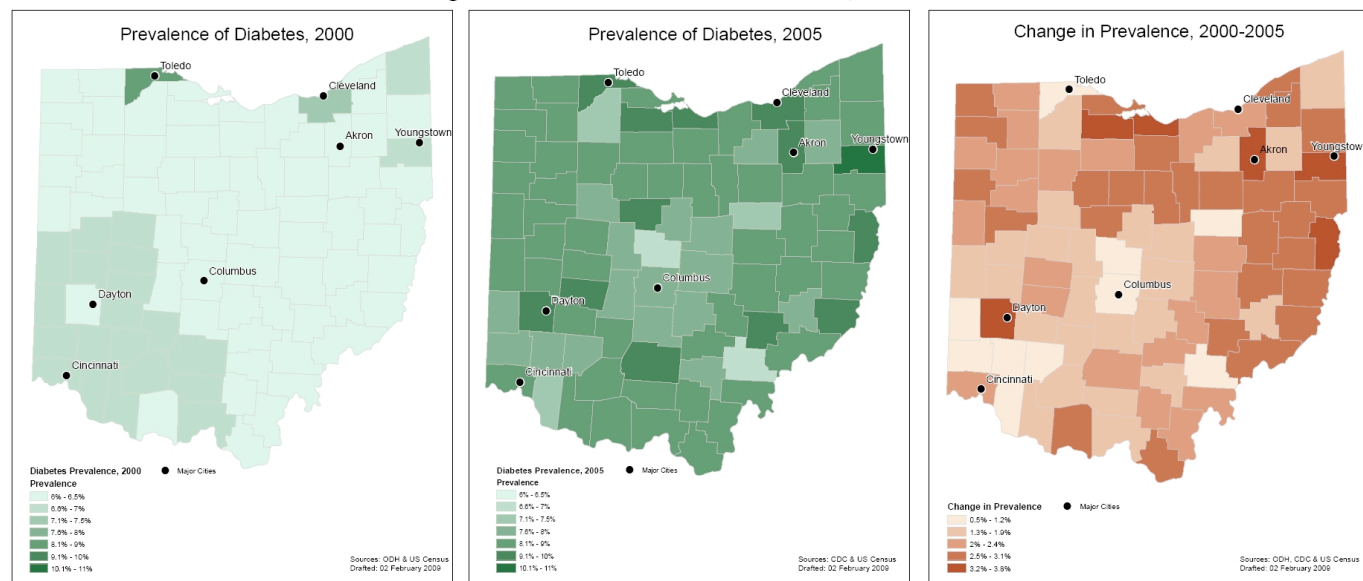


Figure 2- Map of people in poverty (%) per county and DM prevalence

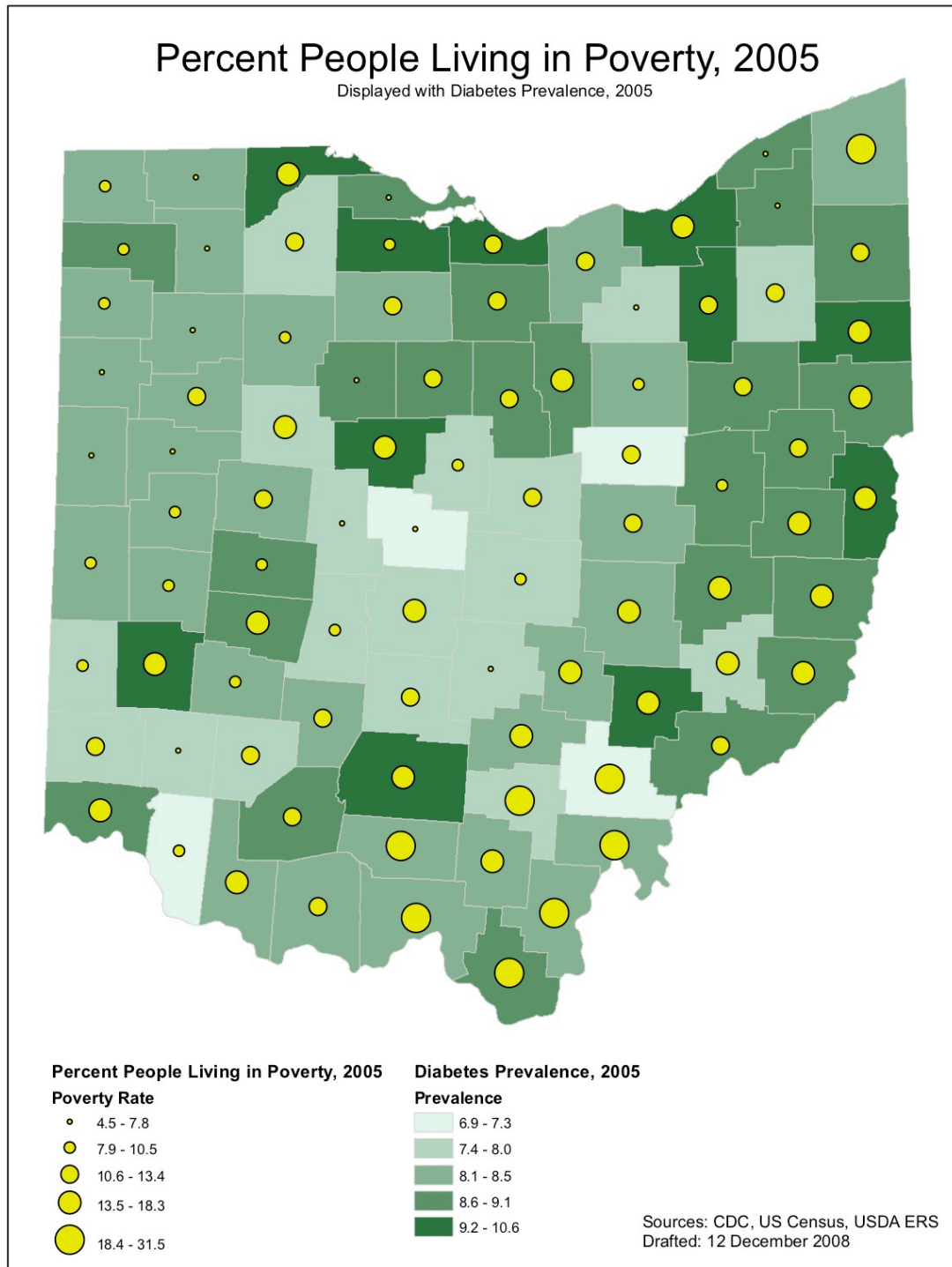


Figure 3- Map of location of DM-related health care providers and prevalence

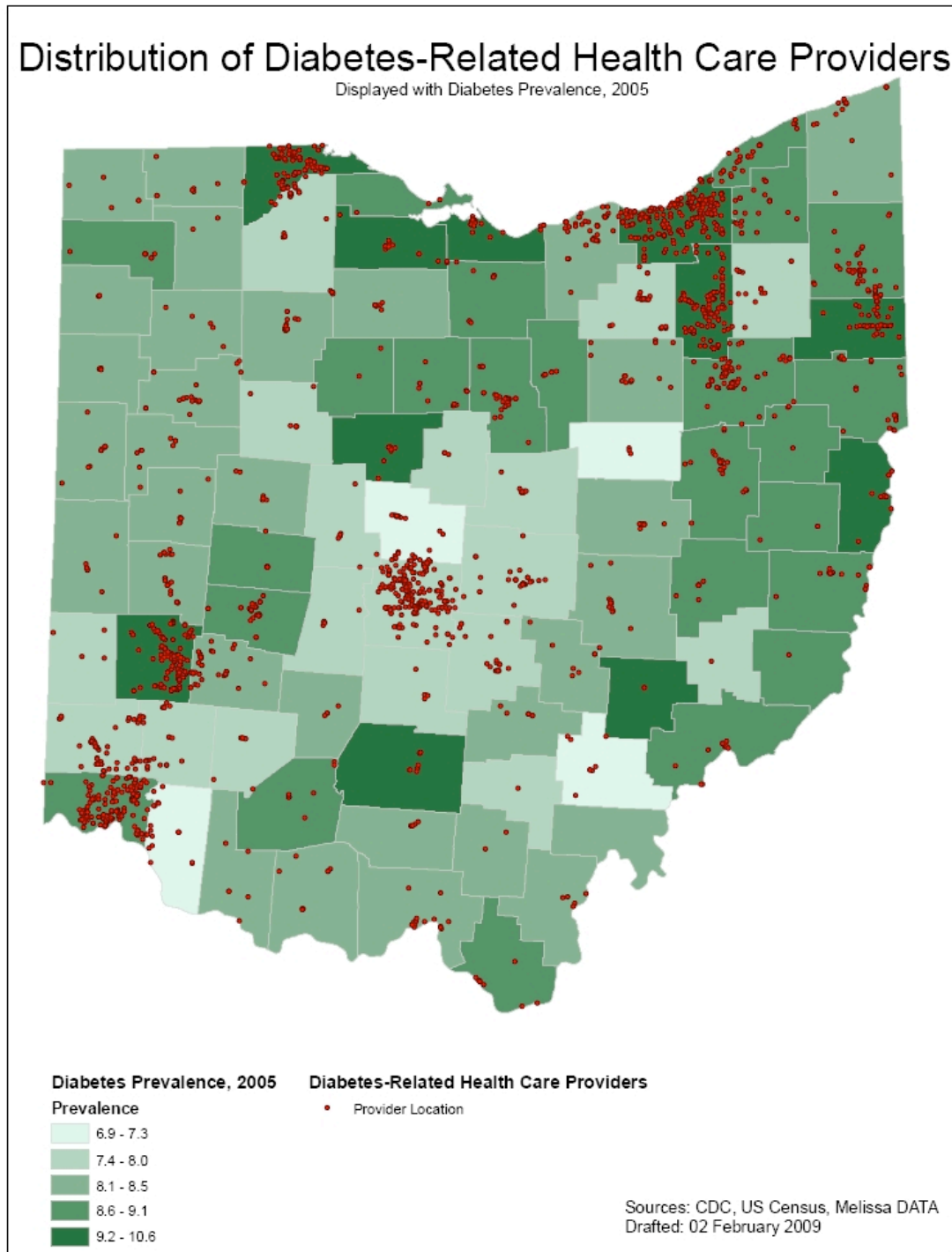
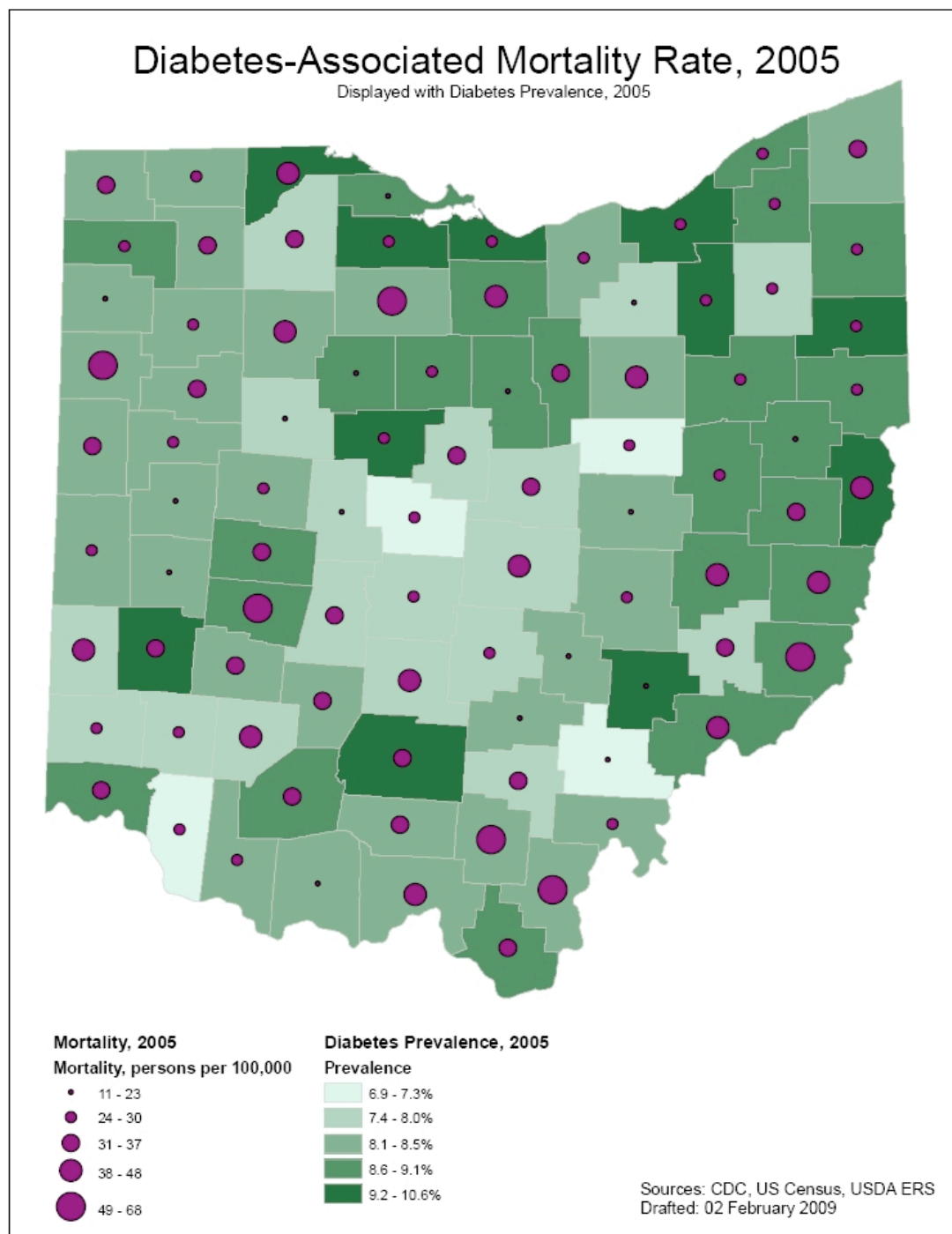


Figure 4- Map of DM-associated mortality and prevalence



References

Caley, L. (2004). Using geographic information systems to design population-based interventions. *Public Health Nursing*, 21 (6), 547-554.

Centers for Disease Control and Prevention: National Diabetes Surveillance System (2005). *National diabetes fact sheet*. Retrieved February 2, 2009 from <http://www.cdc.gov/diabetes/statistics/index.htm>

Choi, M., Afzal, B. & Sattler, B. (2006). Geographic Information Systems: A new health tool for environmental health assessments. *Public Health Nursing*, 23 (5), 381-391.

Coberley, C., Puckrein, G., Dobbs, A., McGinnis, M. Coberley, S. & Shurney, D. (2007). Effectiveness of disease management programs on improving diabetes care for individuals in health disparate areas. *Disease Management*, June, 10 (3).

Croner, C., Sperling, J. & Broome, F. (1996). Geographic Information Systems (GIS): New perspectives in understanding human health and environmental relationships. *Statistics in Medicine*, 15 (17-18), 1961-1977.

Davidson, M. (2004). Establishing prevention, education and community awareness through a comprehensive diabetes, hypertension and hypercholesterolemia screening program: the Smith Island, Maryland, USA experience. *International Journal of Nursing Practice*, October, 10 (5), 242-246.

Division of Prevention, Ohio Department of Health, Ohio Diabetes Prevention and Control Program. (2004, April). *The Burden of Diabetes in Ohio*. Retrieved October 25, 2007, from <http://www.odh.ohio.gov/odhPrograms/hprp/diabetes/dfacts/facts.aspx>

Gesler, W., Hayes, M., Arcury, T., Skelly, A., Nash, S & Soward, A. (2004). Use of mapping technology in health intervention research. *Nursing Outlook*, 52 (3), 142-146.

Huether, S. & McCance, K. (2004). *Understanding Pathophysiology (3rd ed)*. St. Louis: Mosby.

Moss, M. & Schell, M. (2004). GIS(c): A scientific framework and methodological tool for nursing research. *Advances in Nursing Science*, 27 (2), 150-159.

Pronk, N., Boucher, J., Jeffery, R., Sherwood, N. & Boyle, R. (2004). Reducing the incidence of type 2 diabetes mellitus: a healthcare plan and program implementation perspective. *Disease Management and Health Outcomes*, 12(4), 249-58.

Riner, M.E., Cunningham, C. & Johnson, A. (2004). Public health education and practice using geographic information system technology. *Public Health Nursing*, 21 (1), 57-65.

Signorello, L., Schlundt, D., Cohen, S., Steinwandel, M., Buchowski, M., McLaughlin, J., Hargraves, M. & Blot, W. (2007). Comparing diabetes prevalence between African Americans and Whites of similar socioeconomic status. *American Journal of Public Health*, December, 97 (12), 2260-2267.

Susi, L. & Mascarenhas, A.K. (2002). Using a geographical information system to map the distribution of dentists in Ohio. *Journal of American Dental Association*, May, 133, 636-642.

US Census Bureau (2005). *Income and poverty estimates: Ohio counties*. Retrieved December 12, 2008, from
 from http://www.census.gov/cgi-bin/saipe/saipe.cgi?year=2005&type=county&table=county&submit=States%20%26%20Counties&areas=all&display_data=Display%20Data&state=39

Melissa DATA (n.d.). *Offices and clinics of medical doctors*. Retrieved February 2, 2009, from <http://www.melissadata.com/lookups/sic.asp?sic=8011>